Durham University

Durham Research Online

Deposited in DRO:

15 June 2015

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Wood, L.A. and Kendal, R.L. and Flynn, E.G. (2015) 'Does a peer model's task proficiency influence children's solution choice and innovation?', Journal of experimental child psychology., 139 . pp. 190-202.

Further information on publisher's website:

http://dx.doi.org/10.1016/j.jecp.2015.06.003

Publisher's copyright statement:

NOTICE: this is the author's version of a work that was accepted for publication in Journal of Experimental Child Psychology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Experimental Child Psychology, 139, November 2015, 10.1016/j.jecp.2015.06.003.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.

Durham University Library, Stockton Road, Durham DH1 3LY, United Kingdom
Tel : +44 (0)191 334 3042 | Fax : +44 (0)191 334 2971
https://dro.dur.ac.uk

Does a peer-model's task proficiency influence children's solution choice and innovation?

Lara A. Wood[1], Rachel L. Kendal[2] and Emma G. Flynn[3]

[1]Centre for Coevolution of Biology & Culture, Department of Psychology, Durham University, Durham, DH1 3LE, UK.

E-mail: lanw@st-andrews.ac.uk (corresponding author)

[2]Centre for Coevolution of Biology & Culture, Department of Anthropology, Durham University, Durham, DH1 3LE, UK.

[3]Centre for Coevolution of Biology & Culture, School of Education, Durham University, Durham, DH1 1TA, UK.

DRAFT

Acknowledgements

This work was supported by a Durham Doctoral Fellowship to LAW . We would like to thank the children and schools who participated in the study, Andrew Long for assistance with the design and manufacture of the task, Sarah Bedford for assistance with data collection and inter-rater coding and Cameron Turner for inter-rater coding.

Abstract

The current study investigated whether four- to six-year-old children's task solution choice was influenced by the past-proficiency of familiar peer models and the child's personal prior task experience. Peer past-proficiency was established through behavioural assessments of interactions with novel tasks alongside peer and teacher predictions of each child's proficiency. Based on these assessments, one peer model with high past-proficiency and one age-, sex-, dominance-, and popularity-matched peer model with lower past-proficiency were trained to remove a capsule using alternative solutions from a three-solution artificial-fruit task. Video demonstrations of the models were shown to children after they had either a personal successful interaction or no interaction with the task. Generally, there was not a strong bias towards the high past-proficiency model, perhaps due to a motivation to acquire multiple methods and the salience of other transmission biases. However, there was some evidence of a model-based past-proficiency bias; when the high past-proficiency peer matched the participant's original solution there was increased use of that solution whereas if the high past-proficiency peer demonstrated an alternative solution, participants showed increased use of the alternative social solution and novel solutions. Thus, model proficiency influenced innovation.

DRAFT

Does a peer-model's task proficiency influence children's solution choice and innovation?

Laboratory experiments with unfamiliar models enable a controlled investigation of children's social learning strategies, influencing the circumstances under which they copy. However, it is hugely beneficial to look at children's behaviour 'in the wild' (Flynn & Whiten, 2010) implementing a controlled design in a naturalistic setting, such as with familiar peers in a child's classroom or nursery group (Dean, Kendal, Schapiro, Thierry, & Laland, 2012; Flynn & Whiten, 2012). Such paradigms may also identify moments of innovation, whereby children find solutions that have not been socially demonstrated. The current study implemented an experimental procedure designed to mirror a naturalistic context to better understand children's solution choice and innovation relative to (a) the past-proficiency of a known peer model and (b) their personal experience with a task.

Model Past-Proficiency

When faced with divergent novel information from numerous individuals it is adaptive to have a strategy as to whom to copy (Laland, 2004; Rendell et al., 2011). Children demonstrate such model-based biases in their learning (reviewed by Wood, Flynn & Kendal, 2013b). For example, from infancy to six-years, children consistently copy reliable, over unreliable, models for linguistic labelling (Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Vázquez, Delisle, & Saylor, 2012) and artefact use (Birch, et al., 2008; Zmyj, Buttelmann, Carpenter, & Daum, 2010). Copying a proficient, successful model should increase the chances of personal success. In the current study we use the term past-proficiency to refer to a model's domain-specific ability exhibited in the past. As such we focus on the potential for a model to have a reputation for being skilled within the domain that the model is currently demonstrating, and a corresponding model-based bias to influence an observer's solution choice. We used a novel artefacts to establish proficiency reputations so proficiency referred to *successful interaction with novel artefacts*. The child models either scored high in past-proficiency (hitherto 'High PPM') or lower in past-proficiency (hitherto 'Low PPM') pertaining to the relative degree of exploration or, where appropriate, successful extraction of capsules containing stickers from the series of novel artefacts.

The strength of the current study was the use of familiar peer models, enabling an investigation of children's responses to peers based on their actual abilities rather than staged manipulations from two novel actors. However, this paradigm presents challenges. First, peers will differ in past-proficiency *and* in other characteristics such as age, sex, popularity and dominance and these characteristics could also bias children's solution choice. For example, seven- and eight-year-olds copy the food choices of older rather than younger children at the same school (Brody & Stoneman, 1981) and three-year-olds copy the preferences of same-sex (over different-sex) unfamiliar child models for choices of novel food, clothes, toys and games (Frazier, Gelman, Kaciroti, Russell, & Lumeng, 2011; Shutts, Banaji, & Spelke, 2010). These characteristics may also co-vary with proficiency; with an open-diffusion artificial-fruits task, older, more dominant familiar children were watched more and had more successes than younger, less dominant children (Flynn & Whiten, 2012). The second related issue is that young children may struggle to differentiate the subtle differences in their peers' proficiency. For example, whilst Zmyj et al. (2010) differentiated proficiency through a model placing a shoe on his foot or his hand, the current study asked children to imagine who might be better at a task. If this is challenging, children might select peers based on more salient characteristics such as age and sex. To try and evaluate and minimise these challenges age, sex, popularity and dominance measures of the children were taken and analysed in conjunction with peer ratings. Additionally, for the test phase, models were matched on these characteristics.

DRAFT

Prior Experience

Personal prior experience can influence whether a model will be copied; naïve (no prior experience with the task) children that are presented with demonstrations of the same solution faithfully copy this solution, including the copying of causally irrelevant actions, even when other solutions are available, (Bonawitz, Shafto, Gweon, Goodman, Spelke & Schulz, 2011; Flynn & Whiten, 2008; Hopper, Flynn, Wood, & Whiten, 2010; Horner & Whiten, 2005; Horner, Whiten, Flynn, & de Waal, 2006; McGuigan, Whiten, Flynn & Horner, 2007). However, children who interact with a task before witnessing social demonstrations omit subsequently socially-demonstrated causally irrelevant actions, use multiple solutions and explore and innovate new solutions (Wood, Kendal & Flynn, 2013a).

Innovation is defined as producing behaviour that has not been socially observed, like a novel solution, although this does not mean that social information has not contributed to the novel solution (XXX, under revision). Innovation can lead to multiple solutions that increases one's overall knowledge of the task, as well as potentially providing generalisable knowledge regarding the properties of each solution. Wood et al. (2013a) investigated solution choice in naïve children given one social demonstration and previously successful children given a matching or an alternative demonstration. The current study extended this by presenting children with two models demonstrating different solutions; either two novel solutions, or one matching and one novel solution. Giving children multiple social alternatives allowed for further exploration of children's solution choice and innovation.

Summary

The current study investigated solution choice in relation to the proficiency of peer models and children's prior experience with a task. Four- to six-year-olds were selected as

DRAFT

the point of investigation as during this period children develop important cognitive milestones including inhibitory control, false-belief understanding, executive functions along with increased general intelligence, all of which could affect learning in a peer context (Blair & Razza, 2007). For example, a Theory of Mind is associated with increased helping of a novice peer on a novel task (Flynn, 2010). Further, it is at this age that children within the UK start school, and have regular contact with a group of peers, their classmates, thus allowing peer-based social learning strategies to emerge. Testing within a school also allowed for more complex profiling (perceived proficiency, popularity and dominance measures) of the children from the peers and the teachers that had known the children for at least six months. Finally, this focus mirrors and adds to many current studies with this age group. For example, children of this age range have demonstrated high levels of copying causally relevant and irrelevant actions (e.g. McGuigan et al., 2011) which indicates that if no biases exist, imitation levels should be high..

If children were able to identify the more proficienct peers, we predicted that the children who saw two new solutions, presented by a High PPM and Low PPM, would try both demonstrated solutions but would preferentially copy the solution choice of the High PPM. We also predicted that when the High PPM's solution matched the child's original solution and the Low PPM offered an alternative solution children would be more likely to continue using their original solution and less likely to use the alternative social method or innovate other 'unexperienced' solutions relative to when the Low PPM matched the child's solution and the High PPM offered an alternative solution. In line with Wood et al. (2013a) we predicted that those children with no prior personal information would copy a socially demonstrated solution. Conversely, previously successful children would flexibly use personally acquired as well as socially demonstrated solutions and would show innovation through exploring other potential solutions. As described above, investigating such dynamics

DRAFT

allows the complexity of the real-world to be mirrored within an experimentally-controlled investigation, rather than an individual, discrete bias analysis which has been seen in much previous research.

Method

Participants

One hundred and ten children (59 males), aged four-to-six-years-old (range = 56 to 80 months, $M = 65.52$, $SD = 6.00$), were recruited from four primary school classes in County Durham, UK. The class sizes were as follows: class A = 23 (12 males), class B = 27 (14), class C = 28 (13), and class D = 32 (20). There were no significant differences in the number of boys or girls (Binomial $p > .5$). The children had been in their classes for between eight and nine months. Eight children were used as models, five participants were excluded from the analysis due to experimenter error and the experiment was terminated early for three children as they appeared uncomfortable. The remaining 94 children ranged from 57 to 77 months ($M = 65.53$, $SD = 5.74$). There was no significant difference in the distribution of sex (χ^2 (3, $N = 82$) = 0.33, $p = .96$) or age ($F_{3, 78} = 0.10$, $p = .96$) across the five conditions.

Design

The experiment had three phases and participants were systematically allocated (approximate matching of age and sex) to one of four conditions. The presence or absence of an interaction with the task in phase one was the first independent variable: twenty children were selected at random to have no interaction in phase one (Condition 1), while the remaining children (*N = 74*) interacted with the task (all but 12 found a solution). In phase two, all children watched video demonstrations of models using a solution. The second independent variable was how many of the solutions demonstrated were novel to the child

DRAFT

(one or two). In phase three, all children had up to five interactions with the task. All children in Condition 1 one were necessarily assigned to two novel solutions. The other children were systematically assigned to the remaining three conditions (Two novel solutions, High PPM matched and Low PPM demonstrated a novel solution, or Low PPM matched and High PPM demonstrated a novel solution). The dependent variables were attendance to demonstration, solution choice and irrelevant action reproduction.

Selecting Models

In order to ascertain which children were to be models, all children were assessed through group behavioural observations of their interactions with three novel tasks, such that both the experimenter and peers observed recent peer proficiency. Peers and teachers were also asked to rate children's proficiency on novel tasks (Pellegrini, et al. 2007). Additionally, children and teachers rated peers/pupils dominance and popularity to investigate whether proficiency ratings were confounded by these traits (Flynn & Whiten 2012). Thus model selection was rigorous in triangulating various sources of information regarding individual's prior-proficiency reputation (see Table 1 and further detail in the supplementary material).

The original intention was to select models based on behaviour with the three novel tasks as indicated by 'Task Interaction Scores' (TIS), and peer predictions of proficiency. Children's TIS with the three novel tasks was consistent, demonstrating that children's novel task proficiency was robust. However, children were not consistent in their ratings of their peers over a short time period. Furthermore, other characteristics influenced model choice; children of the same sex as the rater and children who were more popular received more peer selections for proficiency. The influence of age approached significance with older children being selected more often as proficient. Conversely, teacher ratings of proficiency correlated with performance on both of the reward tasks. We, therefore, modified the design such that

DRAFT

behavioural performance (TIS), supported by teaching ratings, was prioritised over peer ratings for the choice of models. Models were age, sex, popularity and dominance matched within each class. Details of model selection are summarised in Table 2 (further details in Supplementary Material).

DRAFT

Table 1 Overview of assessments. Teacher rating traits taken from Freeman et al. (2013).

Trait	Name of assessment	Source	Method of assessment
Proficiency	No-Reward Task	**Experimenter Ratings:** Interactions during 15-20 mins of free play with novel tasks	Behavioural assessment of interaction with 3 tasks, including order (relative to the other children) of first proximity (within 1m and oriented towards the task), interaction (placing their hands on part of the task) and success (removing stickers from the task in the reward tasks) as well as frequencies of proximity, interaction and success using one-zero sampling in 40 30second intervals and number of different types of interactions with the task. This resulted in a Task Interaction Score (TIS)
Proficiency	Easy-Reward Task		
Proficiency	Hard-Reward Task		
Proficiency	Proficiency	**Peers:** Asked to pick up to five classmates from photographs	Asked, ‘Which five children would be really good at getting the sticker out of this box?’
Popularity	Popularity		Asked, ‘If you could take five children to a party, who would you take?’
Dominance	Dominance		Asked, ‘Are there any children who like to tell other children what to do?’
Proficiency	Proficiency	**Teachers:** Asked to place photographs of children into one of five groups (Likert scale) for each of the six adjectives	Inquisitive: Likely to explore this task
Proficiency	Proficiency		Intelligent: Quick and accurate in judging and comprehending this task
Proficiency	Proficiency		Inventive: Likely to engage in an inventive behaviour with this task
Popularity	Popularity		Friends with a significant number of others/a smaller number of more influential individuals
Dominance	Aggressive		Often initiates conflicts with other children and dominates resources
Dominance	Unaggressive		Able to acquire and monopolise resources over other individuals

DRAFT

Table 2: Overview of the eight models, two selected from each class

				Novel Task Ranks (Class Median)			Teacher Scores (Class Median)				Peer Scores (Class Median)		
Class	(PPM)	Sex	Age	No Reward	Difficult Reward	Easy Reward	Proficiency	Popularity	Aggressive Dominance	Non-aggressive Dominance	Peer Proficiency	Peer Popularity	Peer Dominance
A	High	M	65	3 (12.5)	3.5 (10.5)	**14 (12.5)**	14 (9.3)	5 (3.5)	3 (2.5)	4 (3)	11 (8)	6 (5)	2 (4)
	Low	M	64	20 (12.5)	12 (10.5)	**12.5 (12.5)**	9 (9.3)	5 (3.5)	2 (2.5)	5 (3)	10 (8)	7 (5)	1 (4)
B	High	F	59	12 (14)	6 (14)	10 (14)	15 (9)	5 (3)	5 (2.5)	5 (3)	7 (9)	4 (3)	3 (2)
	Low	F	60	22 (14)	23 (14)	11 (14)	10 (9)	3 (3)	4 (2.5)	4 (3)	5 (9)	6 (3)	2 (2)
C	High	M	65	4 (12.5)	2 (12.5)	1 (13)	14 (9.8)	4 (3)	3 (2.5)	4 (2.5)	8 (8)	4 (4)	3 (2)
	Low	M	64	13 (12.5)	12.5 (12.5)	22.5 (13)	10 (9.8)	1 (3)	4 (2.5)	4 (2.50	8 (8)	2 (4)	1 (2)
D	High	F	65	**16 (15)**	2 (15.5)	2 (15.5)	15 (8.3)	3 (3.2)	4 (3.2)	5 (3.3)	18 (8.5)	3 (5)	1 (2)
	Low	F	65	**11 (15)**	25 (15.5)	25 (15.5)	5 (8.3)	2 (3.2)	4 (3.2)	3 (3.3)	4 (8.3)	1 (5)	4 (2)

Note. Age = months.

Three tasks (No-Reward, Difficult-Reward, Easy-Reward) = sum of ranks for Task Success with lower scores corresponding better proficiency

Teacher scores = sum (out of 15) of the Mdn score across teachers; Peer proficiency, Popularity and Dominance

Peer scores = sum of nominations by other children; Text in bold indicates unavoidable anomalies to expected rankings

Apparatus

The Sweep-Drawer-Lever Box (SDLB, see Figure 1) is a puzzle box containing a reward held in place by a series of defences. The SDLB is transparent with an opening at the top where a capsule containing a sticker can be inserted. The capsule falls to an opaque green mid-level platform where one of three spatially separated, and functionally unique, manipulandi can be used to push the capsule from the mid-level to a lower level. These three manipulandi are, (1) a silver sweeper with a red handle that moves the capsule to a hole through which the capsule falls, (2) a silver lever used to push the capsule to a hole causing it to fall, and (3) a blue drawer upon which the capsules sits and by pulling the drawer handle, a gap is produced through which the capsule falls. These solutions can also be used in combination and therefore there are seven possible solutions: Sweep, Drawer, Lever, Sweep-Drawer, Drawer-Lever, Lever-Sweep and Drawer-Sweep-Lever, the latter four of these are termed ‘combination-solutions’. On the lower level the capsule rests behind a black door which can be slid to the side to remove it.

DRAFT

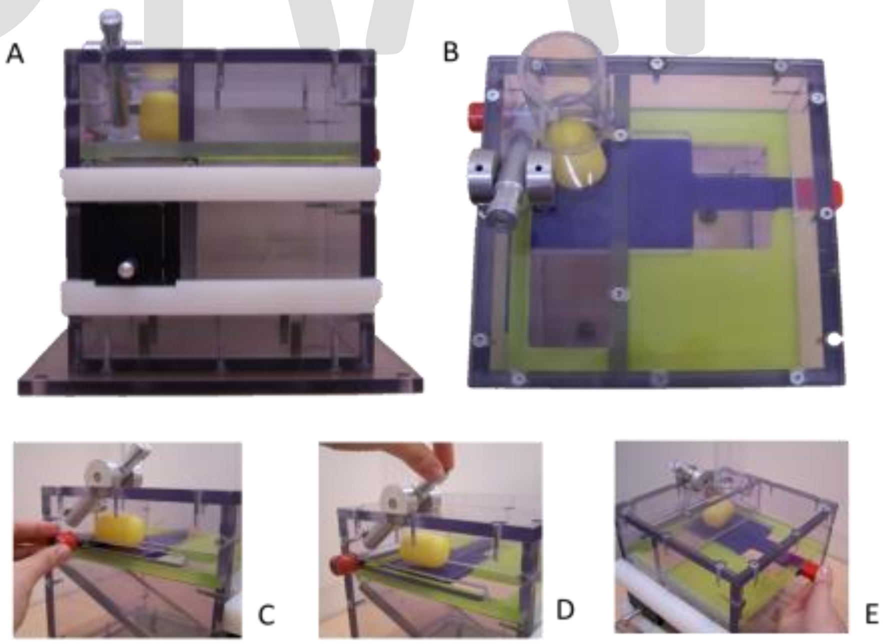

Figure 1. The Sweep-Drawer-Lever Box front view (panel A) and top view (B). Model using the sweep (C), lever (D) drawer (E).

Video Demonstrations

The model demonstrations were presented on two laptops, positioned on a table approximately 30cm apart. Children were initially shown consecutive three-second introductory clips of the models, one model on each laptop, smiling and waving. To aid the child's recall of which model would be shown on which laptop, at the top of each laptop was a photograph (3cm x 5cm) of the corresponding model. Participants were asked to identify each of the models by name. Whether the High PPM or Low PPM model was presented on the left or right and the presentation order of the introductory clips were counterbalanced. The High PPM and Low PPM were individually trained to remove the capsule from the SDLB using the three alternative solutions with each included a sequence of causally irrelevant actions. Once each child was proficient s/he was video recorded completing each of the three sequences of actions ending with the successful extraction of the capsule.

The 15 second clip showed the model looking from the camera to the task, then operating one of the three manipulandi to cause the capsule to fall and then moving this same manipulandi back and forth a further five times (irrelevant actions) before opening the door and retrieving the capsule. For example, if they demonstrated the drawer solution they pulled the drawer out to release the capsule then pushed the drawer (1) in, (2) out, (3) in, (4) out and, (5) in. The model retrieved the capsule from the door and held it up to the camera. Related research has shown a primacy effect such that children preferentially copied the demonstration they saw first (YYY, in prep). Thus the clips were shown simultaneously rather than subsequently so that the participant, not random allocation, dictated who the participant watched first. The clips were shown twice so that, in theory, the child could follow both demonstrations in turn from beginning to end. A video camera was placed 60cm in front of the participant between the two screens. Children's head and eye movements were recorded and coded both for the number of times and total duration of attendance to each demonstration.

DRAFT

Procedure

Children were tested individually in a quiet place in their school. In phase 1 children were given either a chance to interact with the SDLB or were given no information. Children given no information (condition 1 ‘Naïve’) moved straight into phase 2. All other children were assigned to the conditions involving an initial interaction with the SDLB. These children were given three minutes to interact with the task and remove the capsule before moving to phase 2. Children in conditions 1 (‘Naïve’) and 2 (‘Successful’) were presented with novel social information from both models but differed in their prior personal experience. The remaining successful children were given one of two demonstration combinations: a demonstration of the same solution as the child had previously-used presented by the High PPM and an alternative solution by the Low PPM (condition 3 ‘High PPM match’) or a demonstration of the same solution as the child had previously-used by the Low PPM and an alternative solution by the High PPM (condition 4 ‘Low PPM Match’). A summary of all conditions can be found in Table 3. The solution (sweep, lever, drawer) and model type used was counterbalanced across all trials. In phase 3 all children were told, “It’s your turn (again).” The child was allowed to interact with the SDLB until s/he retrieved the capsule successfully or three minutes had elapsed. If children were successful they were told, “It’s your turn again,” until they had completed six trials. At the end of testing all children were told they had done very well and were rewarded with stickers, irrespective of their level of success.

DRAFT

261 *Table 3: Overview of the procedure (three phases) in the five conditions*

	1	2	3	4
Condition	**Naïve**-then-novel-social	**Successful**-then-novel-social	Successful-then-**High PPM-match**-Low-PPM- novel	Successful-then-**Low PPM-match**-High-PPM-novel
Phase 1 (Participant's interaction)	No interaction	Successful	Successful	Successful
Phase 2 (Models' demonstrations)	Two new solutions	Two new solutions	High PPM same solution Low PPM new solution	High PPM new solution Low PPM same solution
Phase 3 (Participant's interaction)	Six Trials	Six Trials	Six Trials	Six Trials
N	20	21	20	21

262 *Note: Words in bold font represent abbreviated terms, used in the text, for condition names.*

DRAFT

Coding, Inter-Rater Reliability and Analysis

Each participant's performance was scored with regard to eye orientation towards each laptop screen during video demonstrations and three separate variables for each response trial: (a) success (capsule removal), (b) solution used, (c) number of causally irrelevant actions copied (out of five). LW coded 100% of the sample from video tape. An independent observer coded 25% of the sample for 22 variables (the three variables listed above for each of six trials and four variables relating to eye orientation). There was almost perfect agreement (Viera & Garrett, 2005) on 21 of the 22 variables (Kappa scores above .86 (p <.01). The remaining variable (the number of causally irrelevant actions on the final trial) had a Kappa score of .64. A second independent observer coded 100% of this variable with a Kappa score of .86 ($p < .01$). All statistical tests were non-parametric and two-tailed.

Results

In phase 1, 62 (84%) of the 74 children who were given a chance to interact with the task were successful. One child used a combination of the Sweep and the Lever action during his success. The other 61 children used a single solution: 19 Drawer, 12 Sweep and 30 Lever. The higher incidence of using the Lever was significant (χ^2 (2, N = 61) = 7.71, $p < .05$). For all subsequent analyses Kruskal-Wallis tests were used to investigate whether the asocial preference for the lever impacted upon results. At no point did the salience of the lever have a significant impact upon the children's subsequent behaviour (all p values > *.05)*. Twelve children were unsuccessful in Phase 1 and so were removed from further analysis.

Children's Attendance to the Demonstrations

Table 4 gives an overview of looking behaviour and times across the two trials for both models. The majority of children alternated their attendance between the two screens during each demonstration (head direction changes between laptops, $M = 4.6$, $SD = 2.4$).

Table 4: Overview of looking behaviour during demonstrations

	Trial 1		Trial 2	
	High PPM	Low PPM	High PPM	Low PPM
Looked First N (%)	45 (56%)	36 (44%)	42 (52%)	39 (48%)
	Binomial $p = .37$		Binomial $p = .82$	
Looked Exclusively at one model (N)	9	3	11	8
Mdn time (seconds) looked at screen (*IQR*)	9 (8.25)	7.5 (8)	6 (7.5)	7 (8.75)
	Wilcoxon Z = -1.5, $p = .14$		Wilcoxon Z = -0.1, $p = .94$	

Past-proficiency Model-based Bias

Across all conditions there was no significant difference in the number of children who used the High PPM solution ($N = 36$) and children who used the Low PPM solution ($N = 31$, Binomial, $p = .63$). A further 13 children used an alternative solution and two children were unsuccessful. Which models' solution (High PPM or Low PPM) was used in the children's first response trial was entered as the dependent variable into a stepwise binary logistic regression with the fixed factors of (a) which model was attended to first (High or Low PPM), (b) the cumulative duration of attendance to each model in both demonstrations, (c) age and (d) sex of participant. The only significant predictor of which model was copied was which model was attended to first (β= -1.14, $p < .05$). Across all conditions the model attended to first, was significantly more likely to be copied than the other model.

Across all six trials there was no significant difference in the number of times the High PPM solution (*Mdn* = 2.0, *IQR* = 3.3) or Low PPM solution was used (*Mdn* = 2.0, *IQR* = 3.0, *Wilcoxon Z* = -0.02, $p = .98$). The number of times a High PPMs solution was used in T1 to T6 was entered as a dependent variable in a Stepwise Linear regression model along with the same four factors. Again, the only significant predictor of High PPM solution use was which model was attended to first (β= 1.08, $p < .05$) with those children that looked at the High PPM model first using the High PPM solution significantly more than those that looked at the Low PPM first.

Conditions where both solutions were novel. There was no significant difference in whether the High or Low PPM's solution was used in the first trial for children in the Naïve condition (High = 7, Low = 11, $p = .48$, Binomial) or Successful condition (High = 8, Low = 7, $p > .99$, Binomial). There was no significant difference in the number of times the children use the High and the Low PPM's solution across all six trials for children in the Naïve condition (High; *Mdn* = 1.0, IQR = 2.8, Low: *Mdn* = 3.0, IQR = 5.0, Wilcoxon Z = -1.48, $p = .14$) or Successful condition (High; *Mdn* = 1.0, IQR = 2.0, Low: *Mdn* = 2.0, IQR = 3.0, Wilcoxon Z = -1.07, *p = .28,*).

Conditions where one model matched the child's personally-acquired solution. Twenty children witnessed the Low PPM match their solution choice and a High PPM demonstrate a new solution. In T1 these children were more likely to use a solution different to their originally discovered (and Low PPM matching) solution ($N = 15$) than use their original solution ($N = 5$, Binomial, $p < .05$). Nine of these 15 children used the High PPM solution and six innovated an *unexperienced* solution. Twenty-one children witnessed the High PPM match their solution and Low PPM demonstrate an alternative. These children were as likely to use their original solution (and High PPM matching) solution ($N = 12$) than

DRAFT

use a solution different to their original (N = 9, Binomial, p = .66) with only one of these children innovating an *unexperienced* solution.

The difference between the two conditions in terms of using their own solution approached significance (p = .06, Fisher's exact two tailed). Across all trials (T1-6), children in the High PPM-same-and-Low PPM-alternate condition continued to use their original solution (Mdn = 5, IQR = 4) more frequently than children in the Low PPM-same-High PPM-alternate condition (Mdn = 2.00, IQR = 6; Z_2= -2.49, p < .05). Not only were children in the Low PPM-same-High PPM-alternate condition more likely to deviate away from their original solution to that demonstrated by the High PPM, but they were significantly more likely to innovate an *unexperienced* solution than children in the High PPM-same-and-Low PPM-alternate condition (χ^2 (1, N = 41) = 7.55, p < .01).

Additional effects of prior experience

All 18 children that had no prior task interaction and were successful (condition 1) used a socially-demonstrated solution in T1. Of the 21 children who discovered a solution in condition 2 (where both models subsequently provide novel solutions), 15 (71%) used a socially-demonstrated solution in T1, so they were significantly more likely to use a socially demonstrated than personally-discovered solution (N = 21, p < .001, Binomial test). However, as six (29%) children did not use a socially demonstrated method, children with prior success were significantly less likely to use a socially-demonstrated solution than children with no prior interaction with a task (Fisher's Exact Test, two tailed, p < .05). Of these six children, two used the solution they had initially discovered and four explored and innovated an *unexperienced* combination-solution (using the same manipulandum used in their personal success, with two adding the Low PPM's manipulandum and two the High PPM's manipulandum). Across the six trials, six of 20 children in the naïve condition used

DRAFT

solutions beyond those experienced at some point. This was not significantly different from the seven of 21 children in the successful condition that discovered multiple solutions (Fisher's Exact test, two tailed, $p = .99$) although these groups are not directly comparable as children in the successful condition had less potential solutions to discover.

Concerning irrelevant action reproduction, the baseline for the rate of spontaneous irrelevant action production was 19% of children. Across all conditions, after social information containing the demonstration of irrelevant actions, there was no increase in the proportion of children producing an irrelevant action (ps > .05). Irrelevant actions were not investigated further.

Discussion

Past-proficiency Model-based Bias

Our prediction that all children would preferentially copy the solution choice of the High PPM over the Low PPM was only partly supported. Children who witnessed two new solutions from differing models did not preferentially copy the solution choice of the High PPM. A null result should be interpreted with caution, especially with a sample size of 20 or 21 per condition, although there was still a null result when the solution choice of all 82 participants were analysed together. Whilst this null result stands in contrast to other studies where children have shown model-based transmission biases for past-proficiency (Birch et al., 2008; Koenig et al., 2004; Koenig & Harris, 2005; Zmyj et al., 2010), our study had a different methodology; (1) both demonstrations offered a viable solution and (2) the models were familiar peers. In relation to difference (1), we suggest that when children observe two new, equally viable solutions they are motivated to try them, irrespective of the source of that useful information. Thus, when both models are effective in the solution they demonstrate their identity is less important; perhaps also explaining why there was no looking preference for the High PPM. These results reflect the complexity of real-world dynamics.

The challenges of using familiar peers had been partly anticipated. Children's ratings of peer proficiency did not correlate with behaviour towards the novel tasks and popularity, age and sex confounded peer ratings. Whilst it could be argued that a failure of the children to identifying proficient peers at this earlier stage undermined the experimental hypothesis and manipulation of peer- proficiency, we think it was fruitful to persevere with the main experiment. This failure was not that surprising considering previous research demonstrating the salience of other characteristics and co-varying characteristics (e.g. Flynn & Whiten, 2012). We factored in the potential for this failure, by taking multiple measures of proficiency, and matched models on age, sex, dominance and popularity. With this matching, children showed a past-proficiency model-based bias under certain conditions, indicating that they have the ability to distinguish between models of varying historical ability and use this to guide their behaviour. Children whose previously-discovered solution was subsequently matched by the High PPM were more likely to continue using this original solution and less likely to use the Low PPM's alternative solution or to innovate, relative to children for whom the Low PPM matched their prior solution use and the High PPM offered an alternative. Children appear to be evaluating their own solution in relation to alternative solutions and the characteristics of the models influenced this evaluation. We suggest that when the child and High PPM's solutions match, that solution is established as a 'good solution' and fidelity towards this solution continues over time. This fidelity inhibits the innovation of alternative solutions. Conversely, when the child's previously-discovered solution matched that of the Low PPM but the High PPM provided an alternate solution, the child perceived that his/her solution (and Low PPM's solution) is only one of many ways to interact with this task and so is motivated to try the alternative offered by the High PPM and innovate unexperienced solutions. Here, investigating the interaction of other model-based biases, such as conformity (see van Leeuwen et al. 2013) would be fruitful, as well as investigating how model-based transmission biases hamper innovation.

As children were not consistent in their peer rating the model choice was based on the relatively objective measures of model proficiency, namely their performance on several tasks (visible to peers) and several teacher ratings of proficiency. Whilst the models may have objectively differed on their proficiency, general perceptions of the models (prior to children observing peers behaviour on the novel tasks) varied greatly amongst each child such that a model might be viewed as proficient by some peers but not by others. Future work could consider the role of peer evaluations of proficiency; however, such future work would need to consider that such ratings can be unreliable and children are prone to rating same-sex older children as proficient, irrespective of the child's proficiency, and thus any resulting bias may be related more to age and sex than proficiency itself.

Children's choice of which demonstration they looked at first was positively correlated to the method used on the first trial. Whilst a bias of 'copy the model observed first' might overwrite a proficiency based bias we think this design was imperative for several reasons. First, it was essential that children were shown both models as previous research (e.g. Wood et al., 2012) indicated that a between-subject design, where a child was shown one model *or* another, can overwrite a model-based bias because any useful social information is better than no information and thus children will copy their one model with high fidelity irrespective of that model's identity. Second, when a child sees both models, allowing a child to select which model to watch first provides a level of ecological validity, and allows us to assess model-based biases using a new approach, different to the usual sequential demonstrations. In related research (YYY in prep), when the demonstrations were sequential and experimentally manipulated, children showed a primacy bias to the first

DRAFT

demonstration, which masked the strength of the bias towards a particular model. Whilst the findings in the current study may have demonstrated a primacy bias, it was the child's choice as to who s/he watched first, and thus who a child chooses to look at first is another measure of a proficiency learning bias rather than a confound to the detection of a proficiency bias.

Prior Experience

Children who discovered a solution and subsequently observed new alternate solutions were motivated to try these new socially-demonstrated solutions, but these children showed less solution canalisation to the socially-demonstrated solutions than naïve children. Indeed, they reverted back to using their original solutions and innovated additional combination-solutions. This finding corresponds to findings from a simpler version of the SDLB (without the lever) where successful personal experience prior to receiving a social demonstration increased solution discovery (Wood et al., 2013a). We suggest that prior personal task success encourages task related self-confidence and this reduces canalisation to social information and encourages innovation. Such a phenomenon has been found in adults whereby participant's confidence in their own response predicted the likelihood of them using subsequent social information, such as those with higher self-confidence were less likely to adopt social information (Morgan, Rendell, Ehn, Hoppitt &Laland, 2011). Developing skills that are immediately unnecessary, but may assist in a changing environment, is thought to underpin instances of contra-freeloading where children (Singh, & Query, 1960) and other animals (Jensen, 1963) work for 'earned' rewards even though identical 'free' awards are available (Inglis, Forkman & Lazarus, 1997). Openness to exploration, innovation, and using multiple solutions for a single challenge may partially underpin cumulative culture, which is widely held to be responsible for the success of humanity as a species (Dean, Vale, Laland, Flynn, & Kendal, 2014).

Causally Irrelevant Actions

Generally, children did not imitate the causally irrelevant actions, contrasting with a number of studies showing that children around this age do so (Horner & Whiten, 2005; Lyons, Young, & Keil, 2007; McGuigan et al., 2007). Previous research has shown minimal copying of casually irrelevant actions when the model is a child (Wood et al., 2012) and when the demonstrations were via video rather than live (McGuigan et al., 2007). Viewing two demonstrations simultaneously may have increased cognitive load and thus decreased the precise copying of a model's solution, although children were able to attend to and copy the relevant solutions. This selective imitation of solution but not causally irrelevant actions could imply that children understand what is causally relevant and what is not and, when a copying context is difficult, parse out non-functional, aspects.

Conclusion

Model solution matching and successful prior experience and influenced children's solution choice and innovation, demonstrating the complex nature of children's social learning strategies. Whilst differences were found in solution choice relative to peer past-proficiency, other model-based biases that occur amongst familiar peers may 'overshadow' a past-proficiency bias. Investigating the relative weightings of different biases is, thus, an important avenue of future research. Biases may encourage or inhibit the innovation of new solutions depending on how they correspond with the child's personal information. In sum, understanding of children's social learning benefits from an approach that emphasises the dynamic setting in which it naturally occurs, enabling consideration of personal experience, number of solutions available, model identity and demonstrations witnessed.

References

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*(3), 1018-1034.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, *78*(2), 647-663.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322-330.

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierrey, B., & Laland, K. N. (2012). Idenitfication of the social-cognitive capabilities underlying human cumulative culture. *Science, 335*(6072), 1114-1118.

Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: a comparative perspective. *Biological Reviews*, 89(2), 284-301.

Flynn, E. (2010). Underpinning collaborative learning. In B. Sokol, U. Muller, J. Carpendale, A. Young, & G. Iarocci (Eds.), Self and social regulation: Social interaction and the development of social understanding and executive functions (pp. 312–336). Oxford, UK: Oxford University Press.

Flynn, E., & Whiten, A. (2008). Cultural transmission of tool use in young children: A diffusion chain study. *Social Development, 17*(3), 699-718.

Flynn, E., & Whiten, A. (2010). Studying children's social learning experimentally "in the wild". *Learning & Behavior, 38*(3), 284-296.

Flynn, E., & Whiten, A. (2012). Experimental "microcultures" in young children: Identifying biographic, cognitive, and social predictors of information transmission. *Child Development, 83*(3), 911-925.

Freeman, H. D., Brosnan, S. F., Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Gosling, S. D. (2013). Developing a comprehensive and comparative questionnaire for measuring personality in chimpanzees using a simultaneous top-down/bottom-up design. *American journal of primatology*, *75*(10), 1042-1053.

Hopper, L. M., Flynn, E. G., Wood, L. A., & Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *Journal of experimental child psychology*, *106*(1), 82-97.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition, 8*(3), 164-181.

Horner, V., Whiten, A., Flynn, E., & de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Science*, *103*, 13878–13883.

Inglis, I. R., Forman, B. M., & Lazarus, J. (1997). Free food or earned food? A review and fuzzy model of contrafreeloading. *Animal Behaviour, 3,* 1171-1191.

Jensen, E.D. (1963). Preference for bar pressing over free-loading as a function of number of unrewarded presses *Journal of Experimental Psychology, 65,* 451-454.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*(10), 694-698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*(6), 1261-1277.

Laland, K. N. (2004). Social learning strategies. *Learning & Behavior, 32*(1), 4-14.

DRAFT

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America, 104*(50), 19751-19756.

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3-and 5-year-old children. *Cognitive Development, 22*(3), 353-364.

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1-18.

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2011). The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences, 279*, 653-662

Pellegrini, A. D., Roseth, C. R., Mliner, S., Bohn, C. M., Van Ryzin, M., Vance, N et al. (2007). Social dominance in preschool classrooms. *Journal of Comparative Psychology, 121*(1), 54-64.

Rendell, L., Fogarty, L., Hoppitt, W. J. E., Morgan, T. J. H., Webster, M. M., & Laland, K. N. (2011). Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences, 15*(2), 68-76.

Singh, D., Query, W. T. (1971). Preference for work over ‘freeloading’ in children. *Psychonomic Science, 24*(2)*,* 77-79

Viera, A.J. and Garrett, J.M. (2005). Understanding inter-observer agreement: The Kappa statistic. *Family Medicine, 37(5)*, 360-363.

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2012). Context dependent model-based biases in cultural transmission: Children’s imitation is affected by model age over model knowledge state. *Evolution and Human Behavior, 33*(4), 387-394.

DRAFT

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2013a). Copy me or copy you? The effect of prior experience on social learning. *Cognition, 127*(2), 203-213.

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2013b). Whom do children copy? Model-based biases in social learning. *Developmental Review*, 33(4), 341-356.

van Leeuwen, E. J., & Haun, D. (2013). Conformity in nonhuman primates: fad or fact?. *Evolution and Human Behavior*, *34*(1), 1-7.

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology, 106*(4), 208-220.

XXX (under revision). Eureka!: What is innovation, how does it develop and who does it? *Child Development*

YYY (in prep). Model-based social learning biases are affected by demonstration content

DRAFT

Supplementary Material Detailing Model Selection

Method

Participants

All 110 children took part in some aspect of peer ratings although there were inevitable absences on some days. Ten members of staff across the four classes assisted with the study. They were all female and worked full time with the respective classes either as the class teacher or as a teaching assistant, all are henceforth described as teachers.

Apparatus

Three tasks were used to assess children’s novel task proficiency (see Figure 1 for details).

DRAFT

Figure 1a. Easy-Reward Task; 150cm long white Perspex pipe with a large hole (d = 10cm) at one end and 12 small holes (d = 3cm) along the pipe. The pipe was filled with shredded paper and approximately 100 stickers removable from the holes. This task was designed to be easy with potentially all stickers being accessed in a 20 minute session.

Figure 1b. *Difficult-Reward Task; 25cm x 25cm x 5cm (h x w x d) transparent Perspex box with six compartments containing shredded paper and around 100 stickers. Each compartment had a hole (d = 3cm) at the front. On the front of the box was a circular transparent Perspex disk (d = 25cm) with four holes (d = 3cm). This circular panel could rotate, allowing the panel hole and the compartment hole to line up for access to the stickers. Two plastic tweezers were attached by a 30cm length of flexible wire and could be used to obtain the stickers. This task was designed to be challenging with potentially only a few stickers being accessed in a 20 minute session.*

Figure 1c. *No-Reward Task; 100cm long transparent Perspex pipe filled with twelve balls of differing colours, sizes and textures. The pipe had three long slats (l= 10cm, w= 2cm) so that children could touch the balls but the balls could not be removed from the pipe.*

DRAFT

Behavioural Proficiency: Assessment and Results

At the beginning of testing the children were told that new toys would be available in 'free-play' and all children could interact with these or they could also choose a different activity. Each novel task was made available to the whole class during these c20min free-play sessions. They were also told that cameras would be recording them, one video camera was placed 2 metres behind the task and another was placed 1 metre to the side of the task. Children's order (relative to the other children) of first proximity (within 1m and oriented towards the task), interaction (placing their hands on part of the task) and success (removing a sticker from the task in the reward tasks) were recorded. Additionally, the frequencies of proximity, interaction and success were recorded using one-zero sampling, whereby the occurrence or absence of each behaviour was noted within 30second intervals. Scores are expressed as the proportion of the 40 potential 30second intervals that a child was in proximity to, interacting or succeeding with the task. Finally, children were scored for the number of different types of interactions with the task (e.g., for the No-Reward task a child could touch the task, insert finger into slots, move ball with finger, move whole tube, interact with the lid, and interact with the zip-ties attaching the task to a rack) and number of stickers obtained (excluding the No-Reward task). Scores were summarised as a 'Task Interaction Score' (TIS).

Pearson rank correlations for behaviour with each of the novel tasks demonstrated that each child's behaviour was similar across the three tasks. No-Reward TIS was positively correlated with Difficult-Reward TIS ($r_{99} = .42$, $p < .001$) and Easy-Reward TIS ($r_{101} = .26$, $p < .01$), which also positively correlated with the Difficult-Reward TIS ($r_{103} = 0.60$, p < .001). On occasion, children were absent during the presentation of one of the novel tasks so the TIS for each task was kept separate. The TIS for each task was entered separately as dependent variables into a stepwise linear regression along with the child's sex (male = 0 or female =1) and age (in months). Sex and age were not significant predictors of the No-Reward TIS. For the Difficult-Reward task age (but not sex) was a significant predictor (β = -0.27, $t_{101} = -2.12$, $p < .05$) of TIS with older children receiving better TIS. For the Easy-Reward task both age (β = -0.30, $t_{103} = -2.71$, $p < .01$) and sex (β = -6.66, $t_{103} = -4.87$, $p < .001$) were significant predictors of TIS with older children performing better than younger children and females having better TIS. To summarise, children showed behavioural consistency across the three tasks and older children and girls tended to have higher TISs, hence demonstrated greater proficiency, than younger children and boys.

Peer Ratings: Assessment and Results

Individually, children were presented with an artificial fruit used in previous social learning research (the transparent version of the Glass Ceiling Box, see Horner & Whiten, 2005). In this task the causally irrelevant actions typically presented with this task were excluded. Children were given a single demonstration of how to retrieve a sticker (by lifting a door, inserting a Velcro topped stick and attaching it to a Velcro sticker) by the experimenter. Children were told it would be their turn after they had answered some questions about their classmates. On a table in front of the participants were photographs of all their classmates, and children were asked three questions; one relating to peer proficiency, "Which five children would be really good at getting the sticker out of this box?" one relating to peer popularity, "If you could take five children to a party with you, who would you take?", one relating to peer dominance, "Are there any children who like to tell other children what to do?" For the last question children struggled to pick five, therefore the question was adapted so children picked up to five peers. The children were then asked again, "Do you remember that I asked which five children would be really good at getting the sticker out of this box? Can you pick those five children again?" This repetition was to ascertain whether responses were consistent over a short amount of time. For each question, the experimenter noted the identity of the five children and then shuffled the photos and randomly distributed them across the table before the next question was asked. At the end of the questioning children were invited to interact with the GCB and were then given a sticker. This interaction served as a means of rewarding children for their participation.

Five children were absent on the day of ratings. Of the 105 children who responded 42 (40%) failed to be consistent in their assessment of peer proficiency, that is, they did not choose at least three of the same five children when asked the same question. There was an interaction between sex of peer and sex of participant with boys choosing more boys (M = 7.37, SD = 2.14) than girls (M = 4.33, SD = 2.61; t_{103} = 6.55, p < .001) and girls choosing more girls (M = 5.65, SD = 2.58) than boys (M = 2.72, SD = 2.27; t_{103} = -6.18, p < .001).

Children's proficiency score (∑peer selections) were entered into a stepwise linear regression in which sex (male = 0 or female =1), age (in months), popularity (∑peer selections) and dominance (∑ peer selections) were entered as predictors. Popularity was the only significant predictor of proficiency rating (β = 1.11, t_{108} = 6.92, p < .001) with such that children who received more peer selections for party attendance receiving more peer selections for proficiency. Sex (β = -0.15, t_{108} = -1.85, p = .067) and age (β = 0.14, t(108) = 1.74, p = .084) approached significance with males and older children being selected more often as proficient. Dominance was not a significant predictor (β = 0.01, t_{108} = -0.14, p = .99) of proficiency. To summarise, children were not consistent in their choices of proficiency of their peers and tended to rate proficiency based on the more popular children of the same sex as themselves.

Teacher Ratings: Assessment and Results

DRAFT

Teachers were shown the same Glass Ceiling Box as an example of a novel task and, in answer to rating statements, were asked to place photos of the children into one of five groups: 1 (not at all like this child), 2 (not like this child), 3 (neither like nor not like this child), 4 (like this child), and 5 (very like this child). The first statements related to proficiency and required teachers to rank children according to: inquisitive, defined as, Likely to explore this task; intelligent, Quick and accurate in judging and comprehending this task; and inventive, Likely to engage in an inventive behaviour with this task. Teachers were also asked to rank children on popularity (Friends with a significant number of others/a smaller number of more influential individuals), aggressive-dominance (Often initiate conflicts with other children and dominates resources) and unaggressive-dominance (Able to acquire and monopolise resources over other individuals without using aggression). These questions were based on constructs developed by Freeman et al. (2013). The scores of same-class teachers were significantly positively correlated with each other for each trait (Table A) with the exception of some of the ratings from teachers of Class A, possibly due to the smaller size (N = 23) and the inquisitive rating in Class B. As there was good agreement amongst the teachers, children received a mean score for each of the six traits. As the three proficiency adjectives were positively correlated (inquisitive with intelligent; $r_{110} = .44$, $p < .001$, and inventive; $r_{110} = .56$, $p < .001$, inventive with intelligent; $r_{110} = 0.71$, $p < .001$) they were combined into a single construct of proficiency.

Table A: Correlations for teachers' rating of children's traits in each of the four classes

	Inquisitive		Intelligent		Inventive		Popularity		Aggression		Unaggressive Dominance	
	Teacher 2	Teacher 3	Teacher 2	Teacher 3	Teacher 2	Teacher 3	Teacher 2	Teacher 3	Teacher 2	Teacher 3	Teacher 2	Teacher 3
Class A: Teacher 1	0.46^{+}	0.78*	0.59*	0.79*	0.37	0.22	0.49	0.64**	0.61*	0.62	0.49+	0.84**
Class B: Teacher 1	0.27	NA	0.71***	NA	0.81***	NA	0.45*	NA	0.66***	NA	0.60**	NA
Class C: Teacher 1	0.48**	NA	0.74***	NA	0.31	NA	0.72***	NA	0.88***	NA	0.52**	NA
Class D: Teacher 1	0.70***	0.51**	0.79***	0.91***	0.83***	0.74***	0.56**	0.53**	0.63***	0.39*	0.62***	0.58***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$ $^{+}p < .075$

DRAFT

Teachers' mean proficiency ratings were entered as a dependent variable into a stepwise linear regression with the child's sex (male = 0 or female = 1), age (in months), popularity, and aggressive and unaggressive dominance entered as independent variables. Age (β = 0.59, $t_{107} = 0.89$, $p = .38$) and sex (β = 0.11, t $t_{107} = 1.73$, $p = .09$) were not significant predictors of teacher ratings of proficiency. Teacher ratings of popularity (β = 1.22, $t_{107} = .20$, $p < .001$), aggressive (β = -0.91, $t(107) = .26$, $p < .01$) and unaggressive dominance (β = 1.48, $t_{107} = .27$, $p < .001$) were all significant predictors of teacher ratings of proficiency, with increased affiliation and unaggressive dominance scores, and decreased aggressive dominance scores, predicting increased proficiency scores. To summarise, it appears that teachers' proficiency judgements were not influenced by children's age or sex, but corresponded positively with ratings of popularity and unaggressive dominance.

Relation between Peer and Teacher Ratings and Behavioural Proficiency

The three TIS were entered separately as dependent variables into a stepwise linear regression with teacher and peer ratings, sex and age, as predictor variables (Table B). For the Easy-Reward task, the model accounting for the most variance (31.7%) of TIS included teacher's proficiency ratings, sex and age; children with higher teacher's proficiency ratings, girls and older children had greater TIS than those with lower proficiency ratings, boys and younger children respectively. For the Difficult-Reward task, teacher proficiency ratings was the only variable in the best model which accounted for 14.7% of the variance of TIS; children with higher proficiency ratings from teachers had greater TIS than those with lower proficiency ratings. For the No-Reward task, the best model could only account for 6.4% of the variance and showed a peculiar pattern. Greater TIS was predicted by increased ratings of aggressive dominance by teachers and *fewer* peer-selection of proficiency.

Table B: Linear Regression (Stepwise) predicting Task Success on two reward novel tasks

	Easy-Reward				Hard-Reward			
Variables in Equation	B	SE	β	*t*	B	SE	β	*t*
Constant	38.1***	6.74		5.6	22.8***	2.19		10.4
Teacher proficiency ratings	-0.79***	0.21	-0.33	-3.8	-0.9***	0.22	-0.39	-43.3
Age (months)	-5.6***	1.31	-0.36	-4.3				
Sex[a]	-0.2*	0.11	-0.17	-2.0				

[a] *Dichotomous variable Male = 0, Female =1; * p < .05, *** p < .001 (two tailed).*

Model Selection

Children were ranked relative to their TIS and teacher proficiency scores. The High PPM was chosen from children who reached the following criteria: in the top 5 TIS rank in at least two of the novel tasks and ranked in the top five children for teacher proficiency rankings. The Low PPM was chosen from children who reached the following criteria: matched the High PPM in sex and age (within 60 days), bottom ten rank for teacher proficiency ratings, did not come in the top ten TIS rank with any novel task. This was possible for three of the four classes, in the fourth class (Class B) no child reached these criteria so for the High PPM model a child who was ranked in the top five children for proficiency by the teachers, and had a TIS rank of 6 and 12 in two novel tasks was selected, and for the Low PPM model a child who was ranked 15th (out of 27) in teacher proficiency and who met the previously described novel task criteria, was selected. All analyses were run with data from this class included and excluded and the results remained the same. The models were also closely matched for popularity and dominance. With popularity, for all classes there was no more than two peer-selections (of a possible range 0-15) difference between models. With unaggressive dominance, for three classes there was no more than three peer-selections (of a possible range 0-7) difference.

DRAFT

References

Freeman, H. D., Brosnan, S. F., Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Gosling, S. D. (2013). Developing a comprehensive and comparative questionnaire for measuring personality in chimpanzees using a simultaneous top-down/bottom-up design. *American journal of primatology*, *75*(10), 1042-1053.

DRAFT